

The Quantum Relative Entropy as a Rate Function and Information Criteria

Kazuya Okamura*

Research Institute of Mathematical Sciences,
Kyoto University, Kyoto, 606-8502, Japan

February 16, 2012

Abstract

We prove that the quantum relative entropy is a rate function in large deviation principle. Next, we define information criteria for quantum states and estimate the accuracy of the use of them. Most of the results in this paper are essentially based on Hiai-Ohya-Tsukada theorem.

1 Introduction

In recent years there is a high level of interest in quantum information theory among researchers. In particular, they are concentrating their studies on control and estimation of quantum states in many systems in order to put it to practical use.

Quantum estimation theory is one of the areas of quantum information theory, which was established in 1970's [16, 19]. Statistical decision theory and hypothesis testing for quantum systems were mainly studied in the beginning stage of this theory. Recently various methods such as information geometry, asymptotic theory, and Bayesian analysis are studied theoretically and applied to many experiments. Especially, it is not too much to say that quantum estimation theory made the monopoly of optimization of measurement.

Our main interest in the paper is asymptotic theory based on large deviation principle. This is the reason why we use the relative entropy, which plays the role of a rate function in large deviation principle and is a typical quasi-distance between probability distributions. The quantum relative entropy is defined analogously and becomes a quasi-distance between quantum states. In early investigations [35, 37, 36], the possibility that it is a rate function in large deviation principle was studied in the context of variational principle. No one however could answer whether this conjecture is correct in a general setting.

A partial answer is given by a quantum version of Stein's lemma in [18, 25, 24], which is conceptually different from the original Stein's lemma [11] and from that in this paper. By contrast, a quantum version of Sanov's theorem which is widely accepted has not appeared [15]. There is only one proposal given in [6].

We will show that some of classical statistical methods are applicable to quantum systems, owing to the universality of statistics and information theory. As a part of this

*e-mail: kazuqi@kurims.kyoto-u.ac.jp

attempt, we will answer the previous question whether the quantum relative entropy is a rate function. In the concrete, a quantum version of Stein's lemma and of Sanov's theorem are proved in Section 4.

Next in Section 5, quantum model selection will be discussed. We try to apply information criteria for quantum states. Information criteria for probability distributions are usually described in terms of the relative entropy, but we derive those for quantum states from the quantum relative entropy. In preceding investigations [43, 47], information criteria are applied to probability distributions calculated from POVM (positive operator-valued measures). Accordingly, the accuracy of the estimation by the use of information criteria for quantum states has not been discussed.

Mathematical and physical backgrounds in this paper are algebraic quantum theory, measuring processes, and Hiai-Ohya-Tsukada theorem. We begin with algebraic quantum theory before stating main results. Its mathematical basis is the theory of operator algebras to clarify the structures of C^* -algebras and of von Neumann algebras, the spaces of linear functionals on them and so on [7, 38, 39]. C^* -algebras are mainly used in the present paper, which describe observables of the system, and equivalence classes of their $*$ -representations, called sectors, play the essential role of classifiers of macroscopic structures in quantum systems. We describe measuring processes within the framework of algebraic quantum theory [13, 28]. We use the (C^*) -tensor product of the algebra of the system and of an apparatus as the algebra of observables of the composite system. Measuring interactions are then defined as automorphisms on this tensor algebra. After these preparations, we can apply Hiai-Ohya-Tsukada theorem [17, Theorem 3.2]. This theorem is crucial for relating the quantum relative entropy with the relative entropy of probability distributions of measured data. We would like to mention that the methods developed in the paper arise from the context of Large Deviation Strategy [29].

Finally, in Section 6, we will discuss the quantum α -divergence [3] as a quasi-entropy formulated by D. Petz [34]. See [3] for its fundamental properties. We will show that the quantum α -divergence is applicable in the (C^*) -algebraic setting. First, for this purpose, the α -analogue of Hiai-Ohya-Tsukada theorem is proved. Secondly, a quantum version of Chernoff bound is proved, which is different from that in [5, 23, 14, 21]. This is a large deviation type estimate, too. Lastly, we define a Bayesian α -predictive state and prove that this minimizes a risk function constructed of the quantum α -divergence. This result generalizes those in [1, 41, 40].

2 Algebraic Quantum Theory and Measuring Processes

2.1 Algebraic Quantum Theory and Preliminaries

Algebraic quantum theory begins with a C^* -algebra \mathfrak{A} of observables of the system. A C^* -algebra \mathfrak{A} is a $*$ -algebra over \mathbb{C} , i.e., an algebra over \mathbb{C} with the involution $*$: $\mathfrak{A} \rightarrow \mathfrak{A}$ defined by

$$(aA + bB)^* = \bar{a}A^* + \bar{b}B^*, \quad (AB)^* = B^*A^*, \quad A^{**} := (A^*)^* = A,$$

for $A, B \in \mathfrak{A}, a, b \in \mathbb{C}$, where \bar{a} is the complex conjugate of a , and also is a Banach space with the norm $\|\cdot\| : \mathfrak{A} \rightarrow \mathbb{R}_{\geq 0} = \{a \in \mathbb{R} \mid a \geq 0\}$ which satisfies, for $A, B \in \mathfrak{A}, a, b \in \mathbb{C}$,

$$\|AB\| \leq \|A\| \|B\|, \quad \|A^*A\| = \|A\|^2, \quad \|A^*\| = \|A\|.$$

In particular, self-adjoint elements $A = A^*$ of \mathfrak{A} are called observables. C^* -algebras \mathfrak{A} are assumed to be unital, i.e., \mathfrak{A} have the unit 1 in the paper. Next, a state ω on \mathfrak{A} is defined as a normalized positive linear functional:

$$\omega(1) = 1, \quad \omega(A^*A) \geq 0, \quad \omega(aA + bB) = a\omega(A) + b\omega(B),$$

for $A, B \in \mathfrak{A}, a, b \in \mathbb{C}$. We denote by $E_{\mathfrak{A}}$ the set of states on \mathfrak{A} . As is seen by the definition, the concept of a state is nothing but a complex-valued output mapping. The following theorem then holds:

Theorem 1 (GNS construction theorem [7, 38]). *Let \mathfrak{A} be a C^* -algebra and ω a state on \mathfrak{A} . Then, there exist a Hilbert space \mathcal{H}_{ω} with the inner product $\langle \cdot, \cdot \rangle$, a vector $\Omega_{\omega} \in \mathcal{H}_{\omega}$, and a $*$ -homomorphism $\pi_{\omega} : \mathfrak{A} \rightarrow \mathcal{B}(\mathcal{H}_{\omega})$, called a $*$ -representation (a representation, for short) of \mathfrak{A} , such that $\omega(A) = \langle \Omega_{\omega}, \pi_{\omega}(A)\Omega_{\omega} \rangle$. The triplet $(\pi_{\omega}, \mathcal{H}_{\omega}, \Omega_{\omega})$ is called a GNS representation of \mathfrak{A} with respect to ω .*

Furthermore, two GNS representations $(\pi_{\omega,1}, \mathcal{H}_{\omega,1}, \Omega_{\omega,1})$, $(\pi_{\omega,2}, \mathcal{H}_{\omega,2}, \Omega_{\omega,2})$ of \mathfrak{A} with respect to ω are unitarily equivalent, i.e., there exists a unitary operator $U : \mathcal{H}_{\omega,1} \rightarrow \mathcal{H}_{\omega,2}$ such that $\pi_{\omega,2}(A) = U\pi_{\omega,1}(A)U^$ for $A \in \mathfrak{A}$ and $\Omega_{\omega,2} = U\Omega_{\omega,1}$.*

We denote by \mathfrak{A}_+^* the set of positive linear functionals on \mathfrak{A} . The GNS construction theorem holds for each $\phi \in \mathfrak{A}_+^*$. Let \mathcal{H} be a Hilbert space. For each subset \mathcal{M} of $B(\mathcal{H})$, let \mathcal{M}' denote the set of all elements in $B(\mathcal{H})$ commuting with every element in \mathcal{M} . A subalgebra \mathcal{M} of $B(\mathcal{H})$ is called a $*$ -subalgebra, if \mathcal{M} is invariant under the involution. Then, a von Neumann algebra \mathcal{M} on \mathcal{H} is a $*$ -subalgebra of $B(\mathcal{H})$ such that $\mathcal{M} = \mathcal{M}'' := (\mathcal{M}')'$, and a factor is a von Neumann algebra \mathcal{M} with trivial center $\mathfrak{Z}(\mathcal{M}) := \mathcal{M} \cap \mathcal{M}' = \mathbb{C}1$. The center $\mathfrak{Z}(\mathcal{M})$ of the von Neumann algebra \mathcal{M} is a unique maximal abelian $*$ -subalgebra of \mathcal{M} commuting with every element of \mathcal{M} . It is well known that, for a subset \mathcal{J} of $B(\mathcal{H})$ which is invariant under the involution, \mathcal{J}'' is the smallest von Neumann algebra containing \mathcal{J} . It is called the von Neumann algebra generated by \mathcal{J} . A typical example can be given by the von Neumann algebra generated by a GNS representation $(\pi_{\omega}, \mathcal{H}_{\omega}, \Omega_{\omega})$ of \mathfrak{A} with respect to ω . Since the subset $\pi_{\omega}(\mathfrak{A}) = \{\pi_{\omega}(A) | A \in \mathfrak{A}\}$ of $B(\mathcal{H}_{\omega})$ is invariant under the involution, $\pi_{\omega}(\mathfrak{A})''$ is the smallest von Neumann algebra on \mathcal{H}_{ω} generated by $\pi_{\omega}(\mathfrak{A})$.

Let π be a representation of \mathfrak{A} . A state ω on \mathfrak{A} is said to be π -normal if there exists a normal state¹ ρ on $\pi(\mathfrak{A})''$ such that $\omega(A) = \rho(\pi(A))$ for every $A \in \mathfrak{A}$. Two representations π_1, π_2 are quasi-equivalent, denoted by $\pi_1 \approx \pi_2$, if each π_1 -normal state is π_2 -normal and vice versa. As a complement, two representations π_1, π_2 are disjoint, denoted by $\pi_1 \circ \pi_2$, if no π_1 -normal state is π_2 -normal and vice versa. A state ω on a C^* -algebra \mathfrak{A} is called a factor state if the center $\mathfrak{Z}_{\omega}(\mathfrak{A}) := \pi_{\omega}(\mathfrak{A})' \cap \pi_{\omega}(\mathfrak{A})''$ of the corresponding von Neumann algebra $\pi_{\omega}(\mathfrak{A})''$ is trivial. We denote by $F_{\mathfrak{A}}$ the set of factor states on \mathfrak{A} .

Definition 1 ([26]). *A sector of C^* -algebra \mathfrak{A} is defined by a quasi-equivalence class of factor states of \mathfrak{A} .*

A key concept of this paper is the sector defined above. Before mentioning how to use it, we introduce the central measure μ_{ω} of a state ω on \mathfrak{A} . For a positive linear functional ω on \mathfrak{A} , a positive regular Borel measure μ on $E_{\mathfrak{A}}$ is called a barycentric measure of ω , which is called the barycenter of μ and denoted also by $b(\mu)$, if it satisfies

$$\omega(A) = \int \rho(A) d\mu(\rho), \quad A \in \mathfrak{A}.$$

¹A positive linear functional ρ on a von Neumann algebra \mathcal{M} on \mathcal{H} is said to be normal if there exists a positive trace class operator $\sigma \in \mathcal{T}(\mathcal{H})_+$ such that $\rho(A) = \text{Tr}[\sigma A]$ for each $A \in \mathcal{M}$.

It is proved in [7, Theorem 4.1.25] that there is a one-to-one correspondence between the abelian von Neumann subalgebras \mathcal{B} of $\pi_\omega(\mathfrak{A})'$ and the orthogonal measures² $\mu =: \mu_{\mathcal{B}}$ of ω . In the case that \mathcal{B} is a subalgebra of the center $\mathfrak{Z}_\omega(\mathfrak{A})$, the corresponding orthogonal measure $\mu_{\mathcal{B}}$ is called a subcentral measure of ω and satisfies the following property: for each $\Delta \in \mathcal{B}(E_{\mathfrak{A}})$, GNS representations π_{ω_Δ} and $\pi_{\omega_{E_{\mathfrak{A}} \setminus \Delta}}$ of $\omega_\Delta = \int_{\Delta} \rho d\mu_{\mathcal{B}}(\rho)$ and $\omega_{E_{\mathfrak{A}} \setminus \Delta} = \int_{E_{\mathfrak{A}} \setminus \Delta} \rho d\mu_{\mathcal{B}}(\rho)$, respectively, are unitarily equivalent to the subrepresentations of π_ω , and are disjoint each other. In particular, $\mu_\omega := \mu_{\mathfrak{Z}_\omega(\mathfrak{A})}$ is called the central measure of ω , and pseudosupported³ by the factor states $F_{\mathfrak{A}}$. That is, the central measure μ_ω of a state ω decomposes its barycenter into different sectors. It is also seen that each state ω has a unique central measure μ_ω and is decomposed into sectors by μ_ω . The physical interpretation of a sector will be discussed in the next subsection.

2.2 Composite Systems and Measuring Processes

A measuring process is a vital physical process to characterize quantum systems and access actual situations. It is widely accepted that apparatuses interact with the system under consideration and output values of observables. This fact is formulated here by a composite system of the system and an apparatus.

Let \mathfrak{A} be a C*-algebra of the system under consideration, ω a state on \mathfrak{A} , and $(\pi_\omega, \mathcal{H}_\omega, \Omega_\omega)$ a GNS representation of \mathfrak{A} with respect to ω . We select an observable $A = A^* \in \pi_\omega(\mathfrak{A})''$ to be measured and use $C_0(A) := \{f(A) | f \in C_0(\text{Sp}(A))\}$ as the algebra of an apparatus measuring A , where the spectrum $\text{Sp}(A) := \{\lambda \in \mathbb{C} | A - \lambda 1 \text{ is not invertible}\} (\subseteq \mathbb{R})$ of A . Then, the algebra of the composite system of the system and the apparatus is defined as $\mathfrak{A} \otimes_{\text{m}} C_0(A)$ ⁴. The von Neumann algebra of this algebra in the representation $\pi_\omega \otimes id$ is calculated as

$$\{(\pi_\omega \otimes id)(\mathfrak{A} \otimes_{\text{m}} C_0(A))\}'' = \pi_\omega(\mathfrak{A})'' \otimes \mathcal{A},$$

where $\mathcal{A} = C_0(\text{Sp}(A))''$ is a von Neumann subalgebra of $\mathcal{B}(\mathcal{H}_\omega)$, with center

$$\begin{aligned} \mathfrak{Z}_{\pi_\omega \otimes id}(\mathfrak{A} \otimes_{\text{m}} C_0(A)) &:= \{(\pi_\omega \otimes id)(\mathfrak{A} \otimes_{\text{m}} C_0(A))\}'' \cap \{(\pi_\omega \otimes id)(\mathfrak{A} \otimes_{\text{m}} C_0(A))\}' \\ &= (\pi_\omega(\mathfrak{A})'' \otimes \mathcal{A}) \cap (\pi_\omega(\mathfrak{A})'' \otimes \mathcal{A})' \\ &= (\pi_\omega(\mathfrak{A})'' \otimes \mathcal{A}) \cap (\pi_\omega(\mathfrak{A})' \otimes \mathcal{A}') \\ &= (\pi_\omega(\mathfrak{A})'' \cap \pi_\omega(\mathfrak{A})') \otimes (\mathcal{A} \cap \mathcal{A}') \\ &= \mathfrak{Z}_\omega(\mathfrak{A}) \otimes \mathcal{A}. \end{aligned}$$

Therefore, the algebra of observables of the apparatus is contained in the center of the algebra of the composite system. Assume, from now on, that the state ω of the system is a factor for simplicity.

Next, let us consider the definition of measuring interactions and their physical meaning. In this paper, measuring interactions are defined by automorphisms on the algebra

²See in [7, Section 4.1].

³A measure μ on a Borel space K is said to be pseudosupported by an arbitrary set $A \subseteq K$ if $\mu(B) = 0$ for all Baire sets B such that $B \cap A = \emptyset$.

⁴The injective C*-tensor product $\mathfrak{A}_1 \otimes_{\text{m}} \mathfrak{A}_2 = \mathfrak{A}_1 \otimes_{\text{min}} \mathfrak{A}_2$ of two C*-algebras \mathfrak{A}_1 and \mathfrak{A}_2 is the completion of the algebraic tensor product $\mathfrak{A}_1 \otimes_{\text{alg}} \mathfrak{A}_2$ by the norm $\|\cdot\|_{\text{min}}$ defined by $\|C\|_{\text{min}} = \sup \|(\pi_1 \otimes \pi_2)(C)\|$, $C \in \mathfrak{A}_1 \otimes_{\text{alg}} \mathfrak{A}_2$, where π_1 and π_2 run over all representations of \mathfrak{A}_1 and \mathfrak{A}_2 , respectively. See [38], in detail.

$\pi_\omega(\mathfrak{A})'' \otimes \mathcal{A}$ of the composite system. When we fix a standard form⁵ of the von Neumann algebra $\pi_\omega(\mathfrak{A})'' \otimes \mathcal{A}$, any automorphism α is unitarily implemented, i.e., there exists a unitary operator U such that $\alpha(A) = U^*AU$, $A \in \pi_\omega(\mathfrak{A})'' \otimes \mathcal{A}$. Owing to this result, we can reproduce unitary operators on a Hilbert space as measuring interactions in general theory of measuring processes in Hilbert space formalism. Let α be an automorphism on a von Neumann algebra \mathcal{M} . We define $\alpha^* : E_{\mathcal{M}} \rightarrow E_{\mathcal{M}}$ by $(\alpha^*\omega)(B) = \omega(\alpha(B))$ for $B \in \mathcal{M}$.

The most important and fundamental measuring interaction is that of ideal measurement of A denoted by $\alpha(W)$, which is to be combined with the concept of a neutral position, vital for the description of function of apparatuses [13]. A neutral position, denoted by $m_{\mathcal{A}}$, is the “universal” initial state of the apparatus and corresponds to the macroscopically stable position of the measuring pointer realized when the apparatus is isolated (for the detail mathematical formulation, see [13]). $\alpha(W)$ is an automorphism on $\pi_\omega(\mathfrak{A})'' \otimes \mathcal{A}$ such that

$$\begin{aligned} \omega_{(A, \alpha(W))}(B) &:= \alpha(W)^*(\tilde{\omega} \otimes m_{\mathcal{A}})((\pi_\omega \otimes id)(B)) \\ &= \int (\omega_a \otimes_m \delta_a)(B) P(A \in da || \omega), \end{aligned} \quad (1)$$

for $B \in \mathfrak{A} \otimes_m C_0(A)$, where $P(A \in \Delta || \omega) = \langle \Omega_\omega, E^A(\Delta) \Omega_\omega \rangle$, E^A the spectral measure of A such that $A = \int a dE^A(a)$, $\Delta \in \mathcal{B}(\mathbb{R})$. $\omega_a = \tilde{\omega}_a \circ \pi_\omega$ such that $\tilde{\omega}_a(f(A)) = f(a)$, δ_a the eigenstate on $C_0(A)$ for $a \in Sp(A)$. The interaction $\alpha(W)$ of ideal measurement should be regarded as a trigger of a perfect correlation [32] between the state of the system and of the apparatus.

Then, Eq.(1) can be rewritten as

$$\begin{aligned} \omega_{(A, \alpha(W))}(B) &= \int \rho(B) d\mu_{\omega_{(A, \alpha)}}(\rho), \\ \text{supp } \mu_{\omega_{(A, \alpha(W))}} &= \overline{\{\omega_a \otimes_m \delta_a | a \in Sp(A)\}}, \end{aligned} \quad (2)$$

for $B \in \mathfrak{A} \otimes_m C_0(A)$. This is because the algebra $\mathfrak{A} \otimes_m C_0(A)$ of the composite system has center $\mathbb{C}1 \otimes \mathcal{A}$, and a family of the eigenvalues of an observable A become an index for classifying sectors. This result shows us that the central measure of a state of the composite system after measurement play the role of a probability measure appearing in Born rule [30]. We can generalize the above argument for any automorphism α on $\pi_\omega(\mathfrak{A})'' \otimes \mathcal{A}$, and discuss the optimization of measurement.

$$\begin{aligned} \omega_{(A, \alpha)}(B) &:= \alpha^*(\tilde{\omega} \otimes m_{\mathcal{A}})((\pi_\omega \otimes id)(B)) \\ &= \int \rho(B) d\mu_{\omega_{(A, \alpha)}}(\rho), \\ \text{supp } \mu_{\omega_{(A, \alpha(W))}} &= \overline{\{\omega_{\alpha, a} \otimes_m \delta_a | a \in Sp(A)\}}. \end{aligned} \quad (3)$$

for $B \in \mathfrak{A} \otimes_m C_0(A)$, where, for each $a \in Sp(A)$, $\omega_{\alpha, a}$ depending on α is the state of the system when the value $a \in Sp(A)$ is measured in an apparatus. Therefore, we use the central measure of a state of the composite system after measuring interaction. If ω above

⁵A standard form of a von Neumann algebra \mathcal{M} on a Hilbert space \mathcal{H} is a quadruple $(\mathcal{M}, \mathcal{H}, J, \mathcal{P})$ consisting of \mathcal{M} , \mathcal{H} , a unitary involution J , a self-dual cone \mathcal{P} in \mathcal{H} , which satisfy the followings: (i) $J\mathcal{M}J = \mathcal{M}'$; (ii) $JAJ = A^*$, $A \in \mathfrak{Z}(\mathcal{M})$; (iii) $J\xi = \xi$, $\xi \in \mathcal{P}$; (iv) $AJAJ\mathcal{P} \subset \mathcal{P}$, $A \in \mathcal{M}$. See [39].

is not a factor, we should replace the central measure $\mu_{\omega_{(A,\alpha)}}$ by the subcentral measure of $\omega_{(A,\alpha)}$ corresponding to the abelian von Neumann subalgebra $\mathbb{C}1 \otimes \mathcal{A}$.

We close this section with discussion on the concept of a sector. A sector can be interpreted as a physical origin of macroscopic indicators of classification, and disjointness among different sectors represents a “conditional” stability of macroscopic structures [27, 26]. As a typical example, sectors in the composite system of the system under consideration with an apparatus are classified by the eigenvalues of an observable to be measured, which was already discussed. We would like to emphasize that quantum measurement theory requires searching nontrivial intersectorial structure in each sector of the system. On the other hand, non-factor states of the system correspond to such physical situations that order parameters like temperature are assumed before any measurement. This case is often overlooked.

3 Hiai-Ohya-Tsukada Theorem and its Application

3.1 Hiai-Ohya-Tsukada Theorem

We denote by $M_1(\Omega)$ the space of probability measures on a measurable space (Ω, \mathcal{F}) . We define the relative entropy of the probability measure $\nu \in M_1(\Omega)$ with respect to $\mu \in M_1(\Omega)$ as

$$D(\nu||\mu) = \begin{cases} \int d\nu(\rho) \log \frac{d\nu}{d\mu}(\rho) & (\nu \ll \mu) \\ +\infty & (\text{otherwise}). \end{cases} \quad (4)$$

If there exists a measure σ on Ω such that $\nu, \mu \ll \sigma$, $D(\nu||\mu)$ is also denoted by $D(q||p)$ where $q := \frac{d\nu}{d\sigma}$ and $p := \frac{d\mu}{d\sigma}$.

Let us explain the formulations of relative entropy due to Araki and to Uhlmann [4, 42, 17]. Let $(\mathcal{M}, \mathcal{H}, J, \mathcal{P})$ be a standard form of a von Neumann algebra \mathcal{M} , and φ, ψ be normal states on \mathcal{M} . There exist unique $\Phi, \Psi \in \mathcal{P}$ such that $\varphi(A) = \langle \Phi, A\Phi \rangle$, $\psi(A) = \langle \Psi, A\Psi \rangle$ for all $A \in \mathcal{A}$. The operator $S_{\Phi, \Psi}$ with the domain $Dom(S_{\Phi, \Psi}) = \mathcal{M}\Psi + (1 - s^{\mathcal{M}'}(\Psi))\mathcal{H}$ is defined by

$$S_{\Phi, \Psi}(A\Psi + \Omega) = s^{\mathcal{M}}(\Psi)A^*\Phi, \quad A \in \mathcal{M}, s^{\mathcal{M}'}(\Psi)\Omega = 0,$$

where $s^{\mathcal{M}}(\Psi)$ is \mathcal{M} -support of Ψ , i.e., the minimal projection E in \mathcal{M} such that $(1 - E)\Psi = 0$. $S_{\Phi, \Psi}$ is seen to be a closable operator. Then the relative modular operator $\Delta_{\Phi, \Psi}$ is defined by $\Delta_{\Phi, \Psi} = (S_{\Phi, \Psi})^* \overline{S_{\Phi, \Psi}}$, and let $\Delta_{\Phi, \Psi} = \int_{-\infty}^{\infty} \lambda dE_{\Phi, \Psi}(\lambda)$ be the spectral decomposition of $\Delta_{\Phi, \Psi}$. Araki's relative entropy $S(\varphi||\psi)_{\text{Araki}}$ is defined by

$$S(\varphi||\psi)_{\text{Araki}} = \begin{cases} \int_{-\infty}^{\infty} \log \lambda d\langle \Phi, E_{\Phi, \Psi}(\lambda)\Phi \rangle, \\ \quad (s(\varphi) \leq s(\psi)), \\ +\infty, \quad (\text{otherwise}), \end{cases}$$

where $s(\varphi)$ is the minimal projection E in \mathcal{M} such that $\varphi(1 - E) = 0$ and called the support of φ .

We then define Uhlmann's relative entropy. For two seminorms p and q on a complex linear space \mathcal{L} , the quadratical mean $QM(p, q)$ is defined by

$$QM(p, q)(x) = \sup_{\alpha \in \mathcal{S}(p, q)} \alpha(x, x)^{\frac{1}{2}}, \quad x \in \mathcal{L},$$

where $\mathcal{S}(p, q)$ is the set of all positive hermitian forms α on \mathcal{L} satisfying $|\alpha(x, y)| \leq p(x)q(y)$ for all $x, y \in \mathcal{L}$. A function $[0, 1] \ni t \mapsto p_t$ whose values are seminorms on \mathcal{L} is called a quadratical interpolation from p to q if it satisfies the following conditions: (i) for each $x \in \mathcal{L}$ the function $t \mapsto p_t(x)$ is continuous; (ii) it satisfies the following properties:

$$\begin{aligned} p_t &= QM(p_{t_1}, p_{t_2}), \quad t = \frac{t_1 + t_2}{2}, t_1, t_2 \in [0, 1], \\ p_{\frac{1}{2}} &= QM(p, q), \\ p_{\frac{t}{2}} &= QM(p, p_t), \quad t \in [0, 1], \\ p_{\frac{1+t}{2}} &= QM(p_t, q), \quad t \in [0, 1]. \end{aligned}$$

Furthermore, for each positive hermitian forms α and β there exists a unique function $[0, 1] \ni t \mapsto QF_t(\alpha, \beta)$ whose values are positive hermitian forms on \mathcal{L} such that the function p_t defined by $p_t(x) = QF_t(\alpha, \beta)(x, x)^{\frac{1}{2}}$ for each $x \in \mathcal{L}$ is the quadratical interpolation from $\alpha(x, x)^{\frac{1}{2}}$ to $\beta(x, x)^{\frac{1}{2}}$. The relative entropy functional $S(\alpha\|\beta)(x)$ of α and β is defined by

$$S(\alpha\|\beta)(x) = -\liminf_{t \rightarrow +0} \frac{QF_t(\alpha, \beta)(x, x) - \alpha(x, x)}{t}.$$

Let \mathcal{A} be a $*$ -algebra, and φ, ψ be a positive linear functionals on \mathcal{A} . The Uhlmann's relative entropy $S(\varphi\|\psi)_{\text{Uhlmann}}$ is defined by

$$S(\varphi\|\psi)_{\text{Uhlmann}} = S(\varphi^R\|\psi^L)(1),$$

where φ^R and ψ^L are the positive hermitian forms given by $\varphi^R(A, B) = \varphi(BA^*)$, $\psi^L(A, B) = \psi(A^*B)$.

In [17], the following three important theorems are proved:

Theorem 2. *For any positive linear functionals φ, ψ on a von Neumann algebra \mathcal{M} , $S(\varphi\|\psi)_{\text{Uhlmann}} = S(\varphi\|\psi)_{\text{Araki}}$.*

Therefore, we do not need to distinguish Araki's relative entropy and Uhlmann's one, and call them the quantum relative entropy. They are denoted by $S(\varphi\|\psi)$.

Theorem 3. *Let φ, ψ be positive linear functionals on a C^* -algebra \mathfrak{A} and π be a nondegenerate representation of \mathfrak{A} on a Hilbert space. If there are the normal extensions $\tilde{\varphi}, \tilde{\psi}$ of φ, ψ to $\pi(\mathfrak{A})''$ such that $\varphi(A) = \tilde{\varphi}(\pi(A))$, $\psi(A) = \tilde{\psi}(\pi(A))$, then $S(\tilde{\varphi}\|\tilde{\psi}) = S(\varphi\|\psi)$.*

We take as π a GNS representation $\pi_{\varphi+\psi}$ corresponding to $\varphi + \psi$ in this paper. The following theorem is our main concern in this section.

Theorem 4 (Hiai-Ohya-Tsukada Theorem[17]). *Let φ, ψ be states on \mathfrak{A} with barycentric measures μ, ν , respectively. If there exists a subcentral measure m such that $\mu, \nu \ll m$, then $S(\varphi\|\psi) = D(\mu\|\nu)$.*

By this theorem we can calculate the quantum relative entropy as the measure-theoretical relative entropy.

3.2 An Application of Theorem 4

Let \mathfrak{A} be a C^* -algebra, ω_1, ω_2 states on \mathfrak{A} , and $(\pi_\omega, \mathcal{H}_\omega, \Omega_\omega)$ a GNS representation of \mathfrak{A} with respect to $\omega = \omega_1 + \omega_2$. We define the set of measurements comparing ω_1 with ω_2 as follows:

$$\begin{aligned} M(\omega_1 \prec \omega_2) &= \bigcup_{A=A^* \in \pi_\omega(\mathfrak{A})''} M(A; \omega_1 \prec \omega_2), \\ M(A; \omega_1 \prec \omega_2) &= \{(A, \alpha) \in \pi_\omega(\mathfrak{A})'' \times \text{Aut}(\pi_\omega(\mathfrak{A})'' \otimes \mathcal{A}) | \\ &\quad \mathcal{A} = C_0(A)'', \mu_{\omega_1, (A, \alpha)} \ll \mu_{\omega_2, (A, \alpha)}\}, \end{aligned}$$

where

$$\begin{aligned} \omega_{1, (A, \alpha)} &= \alpha^*(\tilde{\omega}_1 \otimes m_{\mathcal{A}}) \circ (\pi_\omega \otimes id), \\ \omega_{2, (A, \alpha)} &= \alpha^*(\tilde{\omega}_2 \otimes m_{\mathcal{A}}) \circ (\pi_\omega \otimes id), \end{aligned}$$

and the C^* -algebra of the composite system is $\mathfrak{A} \otimes_m C_0(A)$. In order to give the validity of the definition of $M(\omega_1 \prec \omega_2)$, we prepare the next lemma.

Lemma 1. *Let ω_1, ω_2 be states on a C^* -algebra \mathfrak{A} , $\alpha \in \text{Aut}(\mathfrak{A})$. Then, $S(\omega_1 \| \omega_2) = S(\alpha^* \omega_1 \| \alpha^* \omega_2)$.*

Proof. For a CP map Λ on \mathfrak{A} , it holds that

$$S(\Lambda^* \omega_1 \| \Lambda^* \omega_2) \leq S(\omega_1 \| \omega_2),$$

which is a well-known fact [17, 33]. Thus,

$$S(\alpha^* \omega_1 \| \alpha^* \omega_2) \leq S(\omega_1 \| \omega_2).$$

Since α^{-1} is also an automorphism on \mathfrak{A} ,

$$S(\omega_1 \| \omega_2) = S((\alpha^{-1})^* \alpha^* \omega_1 \| (\alpha^{-1})^* \alpha^* \omega_2) \leq S(\alpha^* \omega_1 \| \alpha^* \omega_2).$$

The lemma is proved. □

If $M(\omega_1 \prec \omega_2) \neq \emptyset$, then for a $(A, \alpha) \in M(\omega_1 \prec \omega_2)$,

$$\begin{aligned} \infty &> D(\mu_{\omega_1, (A, \alpha)} \| \mu_{\omega_2, (A, \alpha)}) \\ &= S(\alpha^*(\tilde{\omega}_1 \otimes m_{\mathcal{A}}) \circ (\pi_\omega \otimes id) \| \alpha^*(\tilde{\omega}_2 \otimes m_{\mathcal{A}}) \circ (\pi_\omega \otimes id)) \\ &= S(\alpha^*(\tilde{\omega}_1 \otimes m_{\mathcal{A}}) \| \alpha^*(\tilde{\omega}_2 \otimes m_{\mathcal{A}})) \\ &= S(\tilde{\omega}_1 \otimes m_{\mathcal{A}} \| \tilde{\omega}_2 \otimes m_{\mathcal{A}}) = S(\tilde{\omega}_1 \| \tilde{\omega}_2) = S(\omega_1 \| \omega_2). \end{aligned}$$

The proof of this equality is essentially based on the use of $(A, \alpha) \in M(\omega_1 \prec \omega_2)$, Theorem 2 and Lemma 1. The central measures $\mu_{\omega_1, (A, \alpha)}, \mu_{\omega_2, (A, \alpha)}$ of states ω_1, ω_2 , respectively, is specified by measured data, since Born rule is equivalent to the central measure of the state of the composite system after measurement. We conclude that we can examine the quantum relative entropy $S(\omega_1 \| \omega_2)$ of ω_1 with respect to ω_2 statistically, whenever $M(\omega_1 \prec \omega_2) \neq \emptyset$.

By the way, we may conjecture the following statement:

Conjecture 1. *If $S(\omega_1 \| \omega_2) < \infty$, then $M(\omega_1 \prec \omega_2) \neq \emptyset$.*

This is, of course, not trivial mathematically. The strong condition $S(\omega_1 \| \omega_2) < \infty$ should be regarded as a physically affirmative statement that ω_1 and ω_2 are so similar that the former can be compared with the latter by the quantum relative entropy $S(\omega_1 \| \omega_2)$. Thus it is expected that the quantum relative entropy corresponds to the measure-theoretical relative entropy using specific probability measures. In the rest of the paper, we assume a condition such as $M(\omega_1 \prec \omega_2) \neq \emptyset$.

4 Large Deviation Type Estimate

In this section, two estimates based on large deviation principle are discussed. One of them is Stein's lemma and the other Sanov's theorem. These are regarded as the standard procedure to give the statistical meaning to the relative entropy. We show here that in quantum systems the quantum relative entropy has the same meaning as the measure-theoretical relative entropy has.

4.1 Another Version of Quantum Stein's Lemma

Assume that $M(\omega_1 \prec \omega_2) \neq \emptyset$, and for a $(A, \alpha) \in M(\omega_1 \prec \omega_2)$ the Radon-Nikodym derivative $\frac{d\mu_{\omega_1, (A, \alpha)}}{d\mu_{\omega_2, (A, \alpha)}}$ is strictly positive. We are now ready for hypothesis testing.

Definition 2. A (decision) test \mathcal{T} is a sequence of measurable functions $T_n : (\text{supp } \mu_{\omega_2, (A, \alpha)})^n \rightarrow \{0, 1\}$ with interpretation $\{T_n = 0\} = \{H_0 \text{ is accepted}\}$ and $\{T_n = 1\} = \{H_1 \text{ is accepted}\}$. If $\mathcal{T} = (T)$, we do not distinguish \mathcal{T} and T .

We define the error probabilities

$$\begin{aligned}\alpha_n^{(A, \alpha)}(T_n) &= P_{\omega_1}^{(A, \alpha)}(T_n = 1) = P_{\omega_1}^{(A, \alpha)}(H_0 \text{ is rejected}), \\ \beta_n^{(A, \alpha)}(T_n) &= P_{\omega_2}^{(A, \alpha)}(T_n = 0) = P_{\omega_2}^{(A, \alpha)}(H_1 \text{ is rejected}).\end{aligned}$$

of the first kind and of the second kind, where $P_{\omega_i}^{(A, \alpha)}$ is the countable product probability measure of $\mu_{\omega_i, (A, \alpha)}$ for $i = 1, 2$. Then, we define the log-likelihood ratio

$$X_j = -\log \frac{d\mu_{\omega_1, (A, \alpha)}}{d\mu_{\omega_2, (A, \alpha)}}(\rho_j)$$

for $\tilde{\rho} =$, and the normalized log-likelihood ratio

$$S_n = \frac{1}{n} \sum_{j=1}^n X_j.$$

A test $\mathcal{T} = (\tilde{T}_n : (\text{supp } \mu_{\omega_2, (A, \alpha)})^n \rightarrow \{0, 1\})$ defined by

$$\tilde{T}_n(\rho^n) = \begin{cases} 0, & (S_n \leq \eta_n), \\ 1, & (\text{otherwise}), \end{cases}$$

is called the Neymann-Pearson test, and denoted also by $\mathcal{T}^\eta = (T_n^{\eta_n})$, where $\eta = (\eta_n)$ is a real sequence.

Lemma 2. For any $0 \leq \gamma \leq 1$, there exists a $\eta = (\eta_n)$ such that $\alpha_n^{(A, \alpha)}(\tilde{T}_n^{\eta_n}) \leq \gamma$, and any other test $\mathcal{T} = (T_n)$ satisfying $\alpha_n^{(A, \alpha)}(T_n) \leq \gamma$ must satisfy $\beta_n^{(A, \alpha)}(T_n) \geq \beta_n^{(A, \alpha)}(\tilde{T}_n^{\eta_n})$.

See [9, 20] for proof. It is well known that both of the error probabilities cannot tend to zero simultaneously. Thus we consider the next quantity.

$$\beta_n^{(A, \alpha)}(\varepsilon) = \inf \{ \beta_n^{(A, \alpha)}(T_n) \mid T_n : \text{test}, \alpha_n^{(A, \alpha)}(T_n) < \varepsilon \}$$

This is the infimum of $\beta_n^{(A, \alpha)}(T_n)$ among all tests T_n with $\alpha_n^{(A, \alpha)}(T_n) < \varepsilon$. It is seen by Lemma 2 that a decision test attaining the infimum is the Neymann-Pearson test $\mathcal{T}^\eta = (T_n^{\eta_n})$, where $\eta = (\eta_n)$ is determined by ε . The following theorem is then proved, which is another version of quantum Stein's lemma.

Theorem 5. For any $\varepsilon < 1$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^{(A, \alpha)}(\varepsilon) = -S(\omega_1 \| \omega_2). \quad (5)$$

Proof. By [11, Lemma 3.4.7 and Exercise 3.4.17],

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^{(A, \alpha)}(\varepsilon) = -D(\mu_{\omega_1, (A, \alpha)} \| \mu_{\omega_2, (A, \alpha)}).$$

Then, $D(\mu_{\omega_1, (A, \alpha)} \| \mu_{\omega_2, (A, \alpha)}) = S(\omega_1 \| \omega_2)$. □

This theorem holds for all C*-algebras, that is, for all quantum systems, if two states φ, ψ are comparable each other, i.e., $M(\varphi \prec \psi) \neq \emptyset$ or $M(\psi \prec \varphi) \neq \emptyset$. Therefore, we expect that the asymptotic theory of classical hypothesis testing is applicable to the quantum case.

4.2 A Quantum Version of Sanov's Theorem

Let \mathfrak{A} be a C*-algebra, ω be a state on \mathfrak{A} , $(\pi_\omega, \mathcal{H}_\omega, \Omega_\omega)$ be a GNS representation of \mathfrak{A} with respect to ω . We use $\mathfrak{A} \otimes_m C_0(A)$, where $A = A^* \in \pi_\omega(\mathfrak{A})''$, as the algebra of the composite system of the system and an apparatus measuring A , and $\alpha \in \text{Aut}(\pi_\omega(\mathfrak{A})'' \otimes \mathcal{A})$, $\mathcal{A} = C_0(A)''$, as a measuring interaction.

We denote by $\mathcal{B}^w(M_1(E_{\mathfrak{A} \otimes_m C_0(A)}))$ the Borel σ -field on $M_1(E_{\mathfrak{A} \otimes_m C_0(A)})$ generated by the weak topology ⁶. For any $\tilde{\rho} = (\rho_1, \rho_2, \dots) \in (\text{supp } \mu_{\omega_{(A, \alpha)}})^{\mathbb{N}}$ and $\Delta \in \mathcal{B}(\text{supp } \mu_{\omega_{(A, \alpha)}})$, we define random variables $Y_j(\tilde{\rho}) = \rho_j$ for $j = 1, 2, \dots$, the empirical measures

$$L_n^{(A, \alpha)}(\tilde{\rho}, \Delta) = \frac{1}{n} \sum_{j=1}^n \delta_{Y_j(\tilde{\rho})}(\Delta), \quad (6)$$

and

$$Q_n^{(A, \alpha)}(\Gamma) = P_\omega^{(A, \alpha)}(L_n^{(A, \alpha)} \in \Gamma), \quad (7)$$

for $\Gamma \in \mathcal{B}^w(M_1(E_{\mathfrak{A} \otimes_m C_0(A)}))$ such that $\{L_n^{(A, \alpha)} \in \Gamma\}$ is measurable, where $P_\omega^{(A, \alpha)}$ denotes the countable product measure of $\mu_{\omega_{(A, \alpha)}}$. Then we state the main theorem in this subsection, the earlier version of which has appeared in [29].

Theorem 6. $Q_n^{(A, \alpha)}$ satisfies the large deviation principle with the good rate function $S(b(\cdot) \| \omega_{(A, \alpha)})$:

$$-S(b(\Gamma^o) \| \omega_{(A, \alpha)}) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log Q_n^{(A, \alpha)}(\Gamma) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log Q_n^{(A, \alpha)}(\Gamma) \leq -S(b(\bar{\Gamma}) \| \omega_{(A, \alpha)}) \quad (8)$$

for $\Gamma \in \mathcal{B}^w(M_1(E_{\mathfrak{A} \otimes_m C_0(A)}))$ such that $\{L_n^{(A, \alpha)} \in \Gamma\}$ is measurable, where

$$S(b(\Gamma) \| \omega_{(A, \alpha)}) := \inf \{S(b(\nu) \| \omega_{(A, \alpha)}) | \nu \in \Gamma, \nu \ll \mu_{\omega_{(A, \alpha)}}\}.$$

In the case that, for $\Gamma \in \mathcal{B}^w(M_1(E_{\mathfrak{A} \otimes_m C_0(A)}))$, $\{\nu \in \Gamma^o | \nu \ll \mu_{\omega_{(A, \alpha)}}\}$ and $\{\nu \in \bar{\Gamma} | \nu \ll \mu_{\omega_{(A, \alpha)}}\}$ are empty, $S(b(\Gamma^o) \| \omega_{(A, \alpha)})$ and $S(b(\bar{\Gamma}) \| \omega_{(A, \alpha)})$ are defined as infinity, respectively.

⁶See [10] and [11, p.261].

Proof. This theorem is easily proved by [10, Sanov's Theorem] and Theorem 4. \square

If $\mathfrak{A} \otimes_m C_0(A)$ is separable, then $E_{\mathfrak{A} \otimes_m C_0(A)}$ becomes a compact metric space whose metric is defined by

$$d(\rho_1, \rho_2) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|\rho_1(A_j) - \rho_2(A_j)|}{\|A_j\|},$$

where the set $\{A_j \neq 0 | j \in \mathbb{N}\}$ is a dense subset of $\mathfrak{A} \otimes_m C_0(A)$. In this case, for all $\Gamma \in \mathcal{B}^w(M_1(E_{\mathfrak{A} \otimes_m C_0(A)}))$, $\{L_n^{(A, \alpha)} \in \Gamma\}$ is measurable. It is known that $B^w(M_1(E_{\mathfrak{A} \otimes_m C_0(A)}))$ is equal to $B^{cy}(M_1(E_{\mathfrak{A} \otimes_m C_0(A)}))$, if $\mathfrak{A} \otimes_m C_0(A)$ is separable, where $B^{cy}(M_1(E_{\mathfrak{A} \otimes_m C_0(A)}))$ is defined as follows. We denote by $B(E_{\mathfrak{A} \otimes_m C_0(A)})$ the vector space of all bounded Borel measurable functions on $E_{\mathfrak{A} \otimes_m C_0(A)}$. For $\phi \in B(E_{\mathfrak{A} \otimes_m C_0(A)})$, let $\tau_\phi : M_1(E_{\mathfrak{A} \otimes_m C_0(A)}) \rightarrow \mathbb{R}$ be defined by $\tau_\phi(\nu) = \langle \phi, \nu \rangle = \int \phi d\nu$. We denote by $\mathcal{B}^{cy}(M_1(E_{\mathfrak{A} \otimes_m C_0(A)}))$ the σ -field of cylinder sets on $M_1(E_{\mathfrak{A} \otimes_m C_0(A)})$, i.e., the smallest σ -field that makes all $\{\tau_\phi\}$ measurable ⁷.

When one of the measures μ attaining the infimum of upper or lower bound is an element of $\Omega^{(A, \alpha)} = \{\nu \in E_{\mathfrak{A} \otimes_m C_0(A)} | b(\nu) = \psi_{(A, \alpha)}, \psi \in E_{\mathfrak{A}}, \psi : \pi_\omega\text{-normal}\}$, there exists at least one state ψ such that $S(b(\nu) \| \omega_{(A, \alpha)}) = S(\psi \| \omega)$. We can therefore reach the state of the system by the use of measured data.

Remark 1. *Optimization of measurement is not considered in this subsection. The setting of Sanov's theorem requires only the knowledge of the methods for collecting and accumulating data. As a result, we should use a measurement (A, α) which is analyzed in detail in a quantum system under consideration.*

5 Quantum Model Selection

First, we define a (parametric) model of states.

Definition 3. *A family of states $\{\omega_\theta | \theta \in \Theta\}$ on \mathfrak{A} is called a (statistical) model if it satisfies the following conditions:*

- (i) Θ is a compact subset of \mathbb{R}^d ($d \in \mathbb{N}$):
- (ii) there exist $\theta_1, \dots, \theta_g \in \Theta$ such that, for every $\theta \in \Theta$, ω_θ is π -normal, where $\pi = \pi_{\omega_{\theta_1} + \dots + \omega_{\theta_g}}$.

Assume that each model in this section is factor for simplicity. A predictive state is a quantum version of predictive (probability) distribution and is a function, which is constructed of a model of states, from data into states. In this section, our purpose is to define a quantum version of information criteria in order to choose the best predictive state from many models. We then define the following measurement.

Definition 4. *For $N \in \mathbb{N}$, N different models $\{\omega_{1, \theta_1} | \theta_1 \in \Theta_1\}, \dots, \{\omega_{N, \theta_N} | \theta_N \in \Theta_N\}$ are comparable if they satisfy the following conditions:*

- (i) there exists $\varphi \in E_{\mathfrak{A}}$ such that $M(\varphi \prec \{\omega_{j, \theta_j}\}_{j=1, \dots, N}) \neq \emptyset$;
- (ii) For a $(A, \alpha) \in M(\varphi \prec \{\omega_{j, \theta_j}\}_{j=1, \dots, N})$,
- (ii-1) there exists a subcentral measure m on $E_{\mathfrak{A} \otimes_m C_0(A)}$ such that $\mu_{j, \theta_j}^{(A, \alpha)} \ll m$ ($j = 1, 2, \dots, N$), where $\mu_{j, \theta_j}^{(A, \alpha)}$ is the central measure of $\omega_{j, \theta_j, (A, \alpha)}$ ($j = 1, 2, \dots, N$);
- (ii-2) the support of $\mu_{j, \theta_j}^{(A, \alpha)}$ is independent of $\theta_j \in \Theta_j$.

⁷See [11, p.263]

The triplet $\mathfrak{P} = (\varphi, (A, \alpha), m)$ is called a predictive measurement for $\{\omega_{1,\theta_1}|\theta_1 \in \Theta_1\}, \dots, \{\omega_{N,\theta_N}|\theta_N \in \Theta_N\}$.

It is easily seen that $\{\omega_{1,\theta_1}|\theta_1 \in \Theta_1\}$, $\{\omega_{2,\theta_2}|\theta_2 \in \Theta_2\}$ and $\{\omega_{3,\theta_3}|\theta_3 \in \Theta_3\}$ are not necessarily comparable, even if $\{\omega_{1,\theta_1}|\theta_1 \in \Theta_1\}$ and $\{\omega_{2,\theta_2}|\theta_2 \in \Theta_2\}$ are comparable and so are $\{\omega_{2,\theta_2}|\theta_2 \in \Theta_2\}$ and $\{\omega_{3,\theta_3}|\theta_3 \in \Theta_3\}$.

Let $\{\omega_{1,\theta_1}|\theta_1 \in \Theta_1\}, \dots, \{\omega_{N,\theta_N}|\theta_N \in \Theta_N\}$ be predictively comparable and $\mathfrak{P} = (\varphi, (A, \alpha), m)$ a predictive measurement for them. For $j = 1, \dots, N$,

$$\begin{aligned}\omega_{j,\theta_j,(A,\alpha)} &= \alpha^*(\tilde{\omega}_{j,\theta_j} \otimes m_A) \circ (\pi_\omega \otimes id) \\ &=: \int \rho d\mu_{j,\theta_j}^{(A,\alpha)}(\rho) \\ &= \int \rho \frac{d\mu_{j,\theta_j}^{(A,\alpha)}}{dm}(\rho) dm(\rho),\end{aligned}$$

and put $p_{j,\theta_j}^{(A,\alpha)}(\rho) = \frac{d\mu_{j,\theta_j}^{(A,\alpha)}}{dm}(\rho)$. We denote by $\rho^n = \{\rho_1, \rho_2, \dots, \rho_n\}$ the set of states consisting of n elements. Then we define a maximal likelihood predictive state $\omega_{\hat{\theta}_j(\rho^n)}$, where $\hat{\theta}_j(\rho^n) = \arg\max_{\theta_j \in \Theta_j} \prod_{i=1}^n p_{j,\theta_j}^{(A,\alpha)}(\rho_i)$ for $\rho^n = \{\rho_1, \rho_2, \dots, \rho_n\}$, and a Bayesian Escort predictive state

$$\omega_{j,\pi_j,\beta}^{\rho^n}(B) = \frac{\int \omega_{j,\theta_j}(B) \prod_{i=1}^n p_{j,\theta_j}^{(A,\alpha)}(\rho_i)^\beta \pi_j(\theta_j) d\theta_j}{\int \prod_{i=1}^n p_{j,\theta_j}^{(A,\alpha)}(\rho_i)^\beta \pi_j(\theta_j) d\theta_j},$$

for $j = 1, 2, \dots, N$, where $\pi_j(\theta_j)$ is a probability density on Θ_j and $\beta > 0$. It then holds that

$$\begin{aligned}\omega_{\hat{\theta}_j(\rho^n),(A,\alpha)} &= \int \rho p_{j,\hat{\theta}_j(\rho^n)}^{(A,\alpha)}(\rho) dm(\rho), \\ \omega_{j,\pi_j,\beta}^{\rho^n} &= \int \rho p_{j,\pi_j,\beta}^{\rho^n,(A,\alpha)}(\rho) dm(\rho),\end{aligned}$$

where

$$p_{j,\pi_j,\beta}^{\rho^n,(A,\alpha)}(\rho) = \frac{\int p_{j,\theta_j}^{(A,\alpha)}(\rho) \prod_{i=1}^n p_{j,\theta_j}^{(A,\alpha)}(\rho_i)^\beta \pi_j(\theta_j) d\theta_j}{\int \prod_{i=1}^n p_{j,\theta_j}^{(A,\alpha)}(\rho_i)^\beta \pi_j(\theta_j) d\theta_j}.$$

Futhermore, we can show the following equalities by Theorem 4:

$$\begin{aligned}S(\varphi||\omega_{\hat{\theta}_j(\rho^n)}) &= \int q(\rho) \log \frac{q(\rho)}{p_{j,\hat{\theta}_j(\rho^n)}^{(A,\alpha)}(\rho)} dm(\rho) = D(q||p_{j,\hat{\theta}_j(\rho^n)}^{(A,\alpha)}), \\ S(\varphi||\omega_{j,\pi_j,\beta}^{\rho^n}) &= \int q(\rho) \log \frac{q(\rho)}{p_{j,\pi_j,\beta}^{\rho^n,(A,\alpha)}(\rho)} dm(\rho) = D(q||p_{j,\pi_j,\beta}^{\rho^n,(A,\alpha)}),\end{aligned}$$

where $q(\rho) = \frac{d\mu_{\varphi(A,\alpha)}}{dm}(\rho)$. Thus we can define information criteria for quantum states, owing to discussion on the formulation of that for probability distributions [2, 44, 45, 46]. Information criteria for quantum states, such as

$$\begin{aligned}\text{AIC}_j &= -\frac{1}{n} \sum_{i=1}^n \log p_{j,\hat{\theta}_j(\rho^n)}^{(A,\alpha)}(\rho_i) + \frac{d_j}{n}, \\ \text{WAIC}_j &= -\frac{1}{n} \sum_{i=1}^n \log p_{j,\pi_j,\beta}^{\rho^n,(A,\alpha)}(\rho_i) + \frac{\beta}{n} \mathcal{V}_j^{(A,\alpha)},\end{aligned}$$

are defined for $j = 1, \dots, N$, where $d_j = \dim \Theta_j$, and $0 < \beta < \infty$, the functional variance $\mathcal{V}_j^{(A,\alpha)}$ defined by

$$\mathcal{V}_j^{(A,\alpha)} = \sum_{i=1}^n \left\{ \langle (\log p_{j,\theta_j}^{(A,\alpha)}(\rho_i))^2 \rangle_{\pi_j,\beta}^{\rho^n} - (\langle \log p_{j,\theta_j}^{(A,\alpha)}(\rho_i) \rangle_{\pi_j,\beta}^{\rho^n})^2 \right\}. \quad (9)$$

The a posteriori mean $\langle \cdot \rangle_{\pi_j,\beta}^{\rho^n}$ in Eq.(9) is defined by

$$\langle G(\theta_j) \rangle_{\pi_j,\beta}^{\rho^n} = \frac{\int G(\theta_j) \prod_{i=1}^n p_{j,\theta_j}^{(A,\alpha)}(\rho_i)^\beta \pi_j(\theta_j) d\theta_j}{\int \prod_{i=1}^n p_{j,\theta_j}^{(A,\alpha)}(\rho_i)^\beta \pi_j(\theta_j) d\theta_j}, \quad (10)$$

for a given function $G(\theta_j)$ on Θ_j . We can choose the best model from $\{\omega_{1,\theta_1} | \theta_1 \in \Theta_1\}, \dots, \{\omega_{N,\theta_N} | \theta_N \in \Theta_N\}$ by information criteria in the setting of a predictive measurement for them. Furthermore, it is obvious by the above discussion that the accuracy of the estimation for quantum states are the same as that for probability measures in the classical case. Thus we have demonstrated the validity of the use of information criteria for models of quantum states. On the other hand, we can conclude that what makes applications of information criteria for quantum states difficult is whether measurements which can compare all models of states under consideration exist or not.

6 On the Quantum α -Divergence

We discuss the quantum α -divergence and its application here.

6.1 The α -version of Hiai-Ohya-Tsukada Theorem

We define the α -divergence

$$D^{(\alpha)}(\mu || \nu) = \begin{cases} D(\nu || \mu), & (\alpha = +1), \\ \frac{4}{1-\alpha^2} \left(1 - \int dm \left(\frac{d\mu}{dm} \right)^{\frac{1-\alpha}{2}} \left(\frac{d\nu}{dm} \right)^{\frac{1+\alpha}{2}} \right), & (|\alpha| < 1), \\ D(\mu || \nu), & (\alpha = -1), \end{cases}$$

for probability measures μ, ν on a measurable space (Ω, \mathcal{F}) which are absolutely continuous with respect to a measure m on (Ω, \mathcal{F}) , Uhlmann's α -divergence

$$S^{(\alpha)}(\varphi || \psi)_{\text{Uhlmann}} = \begin{cases} S(\psi || \varphi)_{\text{Uhlmann}}, & (\alpha = +1), \\ \frac{4}{1-\alpha^2} (1 - QF_{\frac{1+\alpha}{2}}(\varphi^R, \psi^L)(1, 1)), & (|\alpha| < 1), \\ S(\varphi || \psi)_{\text{Uhlmann}}, & (\alpha = -1), \end{cases}$$

for states φ, ψ on a $*$ -algebra, and Araki's α -divergence

$$S^{(\alpha)}(\varphi || \psi)_{\text{Araki}} = \begin{cases} S(\psi || \varphi)_{\text{Araki}}, & (\alpha = +1), \\ \frac{4}{1-\alpha^2} (1 - \langle \Psi, \Delta_{\Phi, \Psi}^{1-\frac{1+\alpha}{2}} \Psi \rangle), & (|\alpha| < 1), \\ S(\varphi || \psi)_{\text{Araki}}, & (\alpha = -1), \end{cases}$$

for normal states $\varphi = \langle \Phi, \cdot \Phi \rangle, \psi = \langle \Psi, \cdot \Psi \rangle$ on a σ -finite von Neumann algebra \mathcal{M} , where Φ, Ψ are elements of a self-dual cone of a standard form. It is easily checked that

$$S^{(\alpha)}(\mu \parallel \nu)_{\text{Araki}} = D^{(\alpha)}(\mu \parallel \nu),$$

for $\mu, \nu \ll m$ on a von Neumann algebra $L^\infty(\Omega, m)$.

The following three theorems are the α -analogue of theorems in subsection 3.1. Proofs are given in Appendix A.

Theorem 7. *For any positive linear functionals φ, ψ on a von Neumann algebra \mathcal{M} , $S^{(\alpha)}(\varphi \parallel \psi)_{\text{Uhlmann}} = S^{(\alpha)}(\varphi \parallel \psi)_{\text{Araki}}$.*

By this theorem, it is not necessary to distinguish Araki's α -divergence and Uhlmann's one. We call them the quantum α -divergence, and denote them by $S^{(\alpha)}(\varphi \parallel \psi)$.

Theorem 8. *Let φ, ψ be positive linear functionals on a C^* -algebra \mathfrak{A} and π be a nondegenerate representation of \mathfrak{A} on a Hilbert space. If there are the normal extensions $\tilde{\varphi}, \tilde{\psi}$ of φ, ψ to $\pi(\mathfrak{A})''$ such that $\varphi(A) = \tilde{\varphi}(\pi(A)), \psi(A) = \tilde{\psi}(\pi(A))$, then $S^{(\alpha)}(\varphi \parallel \psi) = S^{(\alpha)}(\tilde{\varphi} \parallel \tilde{\psi})$.*

Theorem 9. *Let φ, ψ be states on \mathfrak{A} with barycentric measures μ, ν , respectively. If there exists a subcentral measure m such that $\mu, \nu \ll m$, then $S^{(\alpha)}(\varphi \parallel \psi) = D^{(\alpha)}(\mu \parallel \nu)$.*

6.2 A Quantum Version of Chernoff Bound

We use the same notation in subsection 3.2.

$$\begin{aligned} M(\omega_1, \omega_2) &= \bigcup_{A=A^* \in \pi_\omega(\mathfrak{A})''} M(A; \omega_1, \omega_2), \\ M(A; \omega_1, \omega_2) &= \{(A, \alpha) \in \pi_\omega(\mathfrak{A})'' \times \text{Aut}(\pi_\omega(\mathfrak{A})'' \otimes \mathcal{A}) \mid \mathcal{A} = C_0(A)'', \\ &\quad \mu_{\omega_1, (A, \alpha)}, \mu_{\omega_2, (A, \alpha)} \ll \exists m : \text{a subcentral measure}\}, \end{aligned}$$

For a $(A, \alpha) \in M(\omega_1, \omega_2)$, we define $\mathcal{R}_n^{(A, \alpha)}(T_n) = \pi_1 \alpha_n^{(A, \alpha)}(T_n) + \pi_2 \beta_n^{(A, \alpha)}(T_n)$, where $\pi_1, \pi_2 \in \mathbb{R}$ such that $\pi_1, \pi_2 > 0, \pi_1 + \pi_2 = 1$.

Theorem 10.

$$\lim_{n \rightarrow \infty} \inf_{T_n} \frac{1}{n} \log \mathcal{R}_n^{(A, \alpha)}(T_n) = \inf_{0 \leq t \leq 1} \log F_t(\varphi, \psi),$$

where $F_t(\varphi, \psi) = QF_t(\varphi^R, \psi^L)(1, 1)$.

Proof. By [23, Theorem 2.1],

$$\lim_{n \rightarrow \infty} \inf_{T_n} \frac{1}{n} \log \mathcal{R}_n^{(A, \alpha)}(T_n) = \inf_{0 \leq t \leq 1} \log \int dm (p_\varphi^{(A, \alpha)})^{1-t} (p_\psi^{(A, \alpha)})^t,$$

where $p_\varphi^{(A, \alpha)} = \frac{d\mu_{\varphi, (A, \alpha)}}{dm}$, $p_\psi^{(A, \alpha)} = \frac{d\mu_{\psi, (A, \alpha)}}{dm}$.

$$\begin{aligned} \int dm (p_\varphi^{(A, \alpha)})^{1-t} (p_\psi^{(A, \alpha)})^t &= QF_t(\varphi_{(A, \alpha)}^R, \psi_{(A, \alpha)}^L)(1 \otimes_m 1, 1 \otimes_m 1) \\ &= QF_t(\varphi^R, \psi^L)(1, 1) = F_t(\varphi, \psi). \end{aligned}$$

□

6.3 Bayesian α -Predictive State

The concept of a Bayesian α -predictive state or a generalied Bayes predictive state studied in earlier papers [1, 8, 40, 41] is a generalization of a Bayes (escort) predictive state which appeared in the previous section. However, we can easily see that it is difficult to apply this concept in a C^* -algebraic setting. Thus we define a class of statistical model.

Definition 5. A family of states $\{\omega_\theta | \theta \in \Theta\}$ parametrized by a compact set Θ in \mathbb{R}^d is called a classical model if it satisfies the following three conditions:

- (i) There is a subcentral measure m on $E_{\mathfrak{A}}$ such that $\mu_{\omega_\theta} \ll m$ for every $\theta \in \Theta$.
- (ii) The set $\left\{ \rho \in E_{\mathfrak{A}} \left| p_\theta(\rho) := \frac{d\mu_{\omega_\theta}(\rho)}{dm}(\rho) > 0 \right. \right\}$ is independent of $\theta \in \Theta$.
- (iii) ω_θ is Bochner integrable.

Let us define a Bayesian α -predictive state.

Definition 6. Let $\{\omega_\theta | \theta \in \Theta\}$ be a classical model, $\pi(\theta)$ be a probability density on Θ , $\rho^n = \{\rho_1, \dots, \rho_n\}$, $-1 \leq \alpha \leq 1$. A state

$$\omega_{\pi, \alpha}^{\rho^n} = \frac{1}{C_{\pi, \alpha}^{\rho^n}} \int \rho p_{\pi, \alpha}(\rho | \rho^n) dm(\rho), \quad (11)$$

is called a Bayesian α -predictive state, where

$$p_{\pi, \alpha}(\rho | \rho^n) = \begin{cases} \left(\int p_\theta(\rho)^{\frac{1-\alpha}{2}} \pi(\theta | \rho^n) d\theta \right)^{\frac{2}{1-\alpha}}, & (\alpha \neq 1) \\ \exp \left(\int \log p_\theta(\rho) \pi(\theta | \rho^n) d\theta \right), & (\alpha = 1) \end{cases}$$

a posteriori probability density

$$\pi(\theta | \rho^n) = \frac{\pi(\theta) \prod_{j=1}^n p_\theta(\rho_j)}{\int \pi(\theta) \prod_{j=1}^n p_\theta(\rho_j) d\theta}, \quad (12)$$

and

$$C_{\pi, \alpha}^{\rho^n} = \int p_{\pi, \alpha}(\rho | \rho^n) dm(\rho).$$

We prove the following theorem to justify the use of a Bayesian α -predictive state in the context of quantum statistical decision theory, which is first given by [1] and generalized by [8, 41, 40].

Theorem 11. For a state-valued function $\rho^n \mapsto \varphi^{\rho^n}$ such that it has a barycentric measure which is absolutely continuous with respect to m , its risk function

$$R^{(\alpha)}(\omega_\theta \| \varphi^\cdot) = \iint S^{(\alpha)}(\omega_\theta \| \varphi^{\rho^n}) \prod_{i=1}^n d\mu_\theta(\rho_i) \pi(\theta) d\theta \quad (13)$$

is minimized at $\omega_{\pi, \alpha}^{\rho^n}$.

Proof. We prove this theorem only when $\alpha \neq \pm 1$, since $\alpha = \pm 1$ can be proved by similar argument.

$$S^{(\alpha)}(\omega_\theta \| \varphi^{\rho^n}) = \frac{4}{1 - \alpha^2} \int \left(1 - p_\theta(\rho)^{\frac{1-\alpha}{2}} q^{\rho^n}(\rho)^{\frac{1+\alpha}{2}} \right) dm(\rho),$$

where $q^{\rho^n} = \frac{d\mu_{\varphi^{\rho^n}}}{dm}$.

$$\begin{aligned}
& R^{(\alpha)}(\varphi_1 \| \omega_\theta) - R^{(\alpha)}(\varphi_2 \| \omega_\theta) \\
&= \iint \left\{ \frac{4}{1-\alpha^2} \int \left(p_\theta(\rho)^{\frac{1-\alpha}{2}} q_2^{\rho^n}(\rho)^{\frac{1+\alpha}{2}} - p_\theta(\rho)^{\frac{1-\alpha}{2}} q_1^{\rho^n}(\rho)^{\frac{1+\alpha}{2}} \right) dm(\rho) \right\} \prod_{i=1}^n d\mu_\theta(\rho_i) \pi(\theta) d\theta \\
&= \frac{4}{1-\alpha^2} \iiint p_\theta(\rho)^{\frac{1-\alpha}{2}} (q_2^{\rho^n}(\rho)^{\frac{1+\alpha}{2}} - q_1^{\rho^n}(\rho)^{\frac{1+\alpha}{2}}) dm(\rho) \prod_{i=1}^n p_\theta(\rho_i) dm(\rho_i) \pi(\theta) d\theta \\
&= \frac{4}{1-\alpha^2} \iint \left(\int p_\theta(\rho)^{\frac{1-\alpha}{2}} \prod_{i=1}^n p_\theta(\rho_i) \pi(\theta) d\theta \right) (q_2^{\rho^n}(\rho)^{\frac{1+\alpha}{2}} - q_1^{\rho^n}(\rho)^{\frac{1+\alpha}{2}}) dm(\rho) \prod_{i=1}^n dm(\rho_i) \\
&= \frac{4}{1-\alpha^2} \int (C_{\pi,\alpha}^{\rho^n})^{\frac{1-\alpha}{2}} \left\{ \int p_{\pi,\alpha}(\rho|\rho^n)^{\frac{1-\alpha}{2}} (q_2^{\rho^n}(\rho)^{\frac{1+\alpha}{2}} - q_1^{\rho^n}(\rho)^{\frac{1+\alpha}{2}}) dm(\rho) \right\} p_\pi(\rho^n) \prod_{i=1}^n dm(\rho_i)
\end{aligned}$$

where $p_\pi(\rho^n) = \int \prod_{i=1}^n p_\theta(\rho_i) \pi(\theta) d\theta$. When $\varphi_2^{\rho^n}$ is equal to $\omega_{\pi,\alpha}^{\rho^n}$,

$$\begin{aligned}
& R^{(\alpha)}(\varphi_1 \| \omega_\theta) - R^{(\alpha)}(\omega_{\pi,\alpha}^{\rho^n} \| \omega_\theta) \\
&= \frac{4}{1-\alpha^2} \int (C_{\pi,\alpha}^{\rho^n})^{\frac{1-\alpha}{2}} \left\{ \int p_{\pi,\alpha}(\rho|\rho^n)^{\frac{1-\alpha}{2}} (p_{\pi,\alpha}(\rho|\rho^n)^{\frac{1+\alpha}{2}} - q_1^{\rho^n}(\rho)^{\frac{1+\alpha}{2}}) dm(\rho) \right\} p_\pi(\rho^n) \prod_{i=1}^n dm(\rho_i) \\
&= \int (C_{\pi,\alpha}^{\rho^n})^{\frac{1-\alpha}{2}} \left\{ \frac{4}{1-\alpha^2} \int (1 - p_{\pi,\alpha}(\rho|\rho^n)^{\frac{1-\alpha}{2}} q_1^{\rho^n}(\rho)^{\frac{1+\alpha}{2}}) dm(\rho) \right\} p_\pi(\rho^n) \prod_{i=1}^n dm(\rho_i) \\
&= \int (C_{\pi,\alpha}^{\rho^n})^{\frac{1-\alpha}{2}} D^{(\alpha)}(p_{\pi,\alpha}(\rho|\rho^n) \| q_1^{\rho^n}(\rho)) p_\pi(\rho^n) \prod_{i=1}^n dm(\rho_i) \\
&= \int (C_{\pi,\alpha}^{\rho^n})^{\frac{1-\alpha}{2}} S^{(\alpha)}(\omega_{\pi,\alpha}^{\rho^n} \| \varphi_1^{\rho^n}) p_\pi(\rho^n) \prod_{i=1}^n dm(\rho_i) \geq 0.
\end{aligned}$$

The last equality holds when $\varphi_1^{\rho^n}$ equals to $\omega_{\pi,\alpha}^{\rho^n}$, since $\varphi_1^{\rho^n}$ is arbitrary. This means

$$R^{(\alpha)}(\varphi \| \omega_\theta) \geq R^{(\alpha)}(\omega_{\pi,\alpha}^{\rho^n} \| \omega_\theta)$$

for any state-valued function φ^{ρ^n} . □

7 Conclusion

We have proved the validity of the use of Hiai-Ohya-Tsukada theorem on the basis of sector theory and measuring processes. As a result, a quantum version of Stein's lemma and of Sanov's theorem are proved, and it turns out that the quantum relative entropy is a rate function in large deviation theory. In addition, we have formulated the measurement which allows us to apply information criteria, and defined information criteria for quantum states. It is shown that their accuracy is the same as that in classical case. However, there is plenty of room for deepening measurements which we can evaluate for each model equivalently.

Let us compare the results in this paper with past studies. The methods in this paper are extensions of classical ones, which suit modern statistics and stand different views from

[16, 19] and succeeding investigations. In particular, model selection using information criteria compares with hypothesis testing as methods for constructing models and testing hypotheses, and is expected to apply to quantum systems effectively. In asymptotic theory of quantum hypothesis testing, we could consider the universal situation increasing the number of measured data in this paper, while the special but important situation that the number of quanta in the system increases were examined in past studies [5, 14, 15, 18, 21, 23, 24, 25]. It is common for both of them that optimization of measurement is vital. We hope that both universal and special methods in quantum statistical inference will be developed in the future.

A Proofs of Theorems

Proof of Theorem 7. It suffices to prove for $|\alpha| < 1$. It is proved by discussion in [17, p.129] that

$$QF_t(\varphi^R, \psi^L)(1, 1) = \langle \Psi, \Delta_{\Phi, \Psi}^{1-t} \Psi \rangle.$$

Therefore, we have

$$\begin{aligned} S^{(\alpha)}(\varphi \parallel \psi)_{\text{Uhlmann}} &= \frac{4}{1 - \alpha^2} (1 - QF_{\frac{1+\alpha}{2}}(\varphi^R, \psi^L)(1, 1)) \\ &= \frac{4}{1 - \alpha^2} (1 - \langle \Psi, \Delta_{\Phi, \Psi}^{\frac{1-\alpha}{2}} \Psi \rangle) \\ &= S^{(\alpha)}(\varphi \parallel \psi)_{\text{Araki}}. \end{aligned}$$

□

Proof of Theorem 8. It suffices to prove for $|\alpha| < 1$. By [17, Lemma 3.1], it is proved that

$$QF_t(\varphi^R, \psi^L)(A, A) = QF_t(\tilde{\varphi}^R, \tilde{\psi}^L)(\pi(A), \pi(A)),$$

where $\pi = \pi_{\varphi+\psi}$. Thus we have

$$\begin{aligned} S^{(\alpha)}(\varphi \parallel \psi) &= \frac{4}{1 - \alpha^2} (1 - QF_{\frac{1+\alpha}{2}}(\varphi^R, \psi^L)(1, 1)) \\ &= \frac{4}{1 - \alpha^2} (1 - QF_{\frac{1+\alpha}{2}}(\tilde{\varphi}^R, \tilde{\psi}^L)(\pi(1), \pi(1))) \\ &= S^{(\alpha)}(\tilde{\varphi} \parallel \tilde{\psi}). \end{aligned}$$

□

Lemma A.1. *Let $\varphi|_{\mathcal{B}}, \psi|_{\mathcal{B}}$ be the restrictions of states φ, ψ on a von Neumann algebra \mathcal{M} to a $*$ -subalgebra \mathcal{B} to \mathcal{M} . Then,*

$$S_{\mathcal{B}}^{(\alpha)}(\varphi \parallel \psi) := S^{(\alpha)}(\varphi|_{\mathcal{B}} \parallel \psi|_{\mathcal{B}}) \leq S^{(\alpha)}(\varphi \parallel \psi).$$

Proof. It suffices to prove it for $|\alpha| < 1$. $|_{\mathcal{B}} : E_{\mathcal{M}} \rightarrow E_{\mathcal{B}}$ is the dual of the embedding $\mathcal{B} \ni B \mapsto B \in \mathcal{M}$. Using [42, Proposition 9] and Theorem 8,

$$\begin{aligned} QF_t(\varphi^R \parallel \psi^L)(1, 1) &= QI_t(\varphi^R \parallel \psi^L)(1)^2 \\ &\leq QI_t((\varphi|_{\mathcal{B}})^R \parallel (\psi|_{\mathcal{B}})^L)(1)^2 \\ &= QF_t((\varphi|_{\mathcal{B}})^R \parallel (\psi|_{\mathcal{B}})^L)(1, 1). \end{aligned}$$

Thus we have

$$\begin{aligned}
S^{(\alpha)}(\varphi\|\psi) &= \frac{4}{1-\alpha^2}(1 - QF_{\frac{1+\alpha}{2}}(\varphi^R, \psi^L)(1, 1)) \\
&\geq \frac{4}{1-\alpha^2}(1 - QF_{\frac{1+\alpha}{2}}((\varphi|_{\mathcal{B}})^R \| (\psi|_{\mathcal{B}})^L)(1, 1)) \\
&= S^{(\alpha)}(\varphi|_{\mathcal{B}}\|\psi|_{\mathcal{B}}) = S_{\mathcal{B}}^{(\alpha)}(\varphi\|\psi).
\end{aligned}$$

□

Proof of Theorem 9. It suffices to prove for $|\alpha| < 1$.

We define $\Gamma : \mathfrak{A} \rightarrow C(E_{\mathfrak{A}})$ by $\Gamma(A)(\omega) = \omega(A)$, and its dual map $\Gamma^* : M_1(E_{\mathfrak{A}}) \rightarrow E_{\mathfrak{A}}$ by

$$(\Gamma^*\lambda)(A) = \int \omega(A) d\lambda(\omega) = \lambda(\Gamma(A)),$$

for $\lambda \in M_1(E_{\mathfrak{A}})$. Γ satisfies

$$\Gamma(1) = 1, \quad \Gamma(A^*) = \Gamma(A)^*, \quad \Gamma(A^*)\Gamma(A) \leq \Gamma(A^*A),$$

for any $A \in \mathfrak{A}$. Thus,

$$S^{(\alpha)}(\varphi\|\psi) = S^{(\alpha)}(\Gamma^*\mu\|\Gamma^*\nu) \leq D^{(\alpha)}(\mu\|\nu).$$

By assumption, m is a subcentral measure $\mu_{\mathcal{C}}$ associated with a von Neumann subalgebra \mathcal{C} of the center $\mathfrak{Z}_{\chi}(\mathfrak{A}) = \pi_{\chi}(\mathfrak{A})'' \cap \pi_{\chi}(\mathfrak{A})'$ of a state $\chi \in E_{\mathfrak{A}}$. We define a $*$ -isomorphism $\kappa_m : L^{\infty}(E_{\mathfrak{A}}, m) \rightarrow \mathcal{C}$ by

$$\langle \Omega_{\chi}, \pi_{\chi}(A) \kappa_m(f) \Omega_{\chi} \rangle = \int f(\rho) \rho(A) dm(\rho).$$

It is then proved in [17, Example 2.6 and Theorem 3.2] that

$$\begin{aligned}
\tilde{\varphi}(\kappa_m(f)) &= \int f(\rho) d\mu(\rho), \\
\tilde{\psi}(\kappa_m(f)) &= \int f(\rho) d\nu(\rho),
\end{aligned}$$

where $\tilde{\varphi}, \tilde{\psi}$ are the normal extensions of φ, ψ to $\pi_{\chi}(\mathfrak{A})''$. By Lemma A.1,

$$\begin{aligned}
S^{(\alpha)}(\varphi\|\psi) &= S^{(\alpha)}(\tilde{\varphi}\|\tilde{\psi}) \\
&\geq S_{\mathcal{C}}^{(\alpha)}(\tilde{\varphi}\|\tilde{\psi}) = D^{(\alpha)}(\mu\|\nu).
\end{aligned}$$

□

Acknowledgment

The author would like to thank Prof. Izumi Ojima, Dr. Hayato Saigo, Dr. Takahiro Hasebe, Dr. Hiroshi Ando for their warm encouragement and useful comments.

References

- [1] J. Aitchison, Goodness of prediction fit, *Biometrika* **62**, 547-554 (1975).
- [2] H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automatic Control* **19**, 716-723 (1974).
- [3] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of mathematical monographs; v. 191, Amer. Math. Soc. & Oxford Univ. Press (2000).
- [4] H. Araki, Relative entropy for states of von Neumann algebras II. *Publ. RIMS, Kyoto Univ.* **13** (1977), 173-192.
- [5] K.M.R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acín, and F. Verstraete, Discriminating states: The quantum Chernoff bound, *Phys. Rev. Lett.* **98**, 160501 (2007).
- [6] I. Bjelakovic, J.-D. Deuschel, T. Kruger, R. Seiler, R. Siegmund-Schultze and A. Szkola, A quantum version of Sanov's theorem, *Commun. Math. Phys.* **260**, (2006), pp.659-671.
- [7] O. Bratteli and D.W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (vol.1) (2nd printing of 2nd ed.), (Springer, 2002).
- [8] J.M. Corcuera and F. Giummolè, A generalized Bayes rule for prediction, *Scand. J. Statist.* **26**, 265 (1999).
- [9] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, (Wiley, New York, 1991).
- [10] I. Csiszár, A simple proof of Sanov's theorem, *Bull. Brazilian Math. Soc.* **37**, (2006) 453-459.
- [11] A. Dembo and O. Zeitouni, *Large deviations techniques and applications* (2nd ed.), (Springer, 2002).
- [12] M. Drton, B. Sturmfels, and S. Sullivan, *Lectures on Algebraic Statistics*, (Birkhäuser, 2008).
- [13] R. Harada and I. Ojima, A unified scheme of measurement and amplification processes based on Micro-Macro Duality –Stern-Gerlach experiment as a typical example–, *Open Sys. Inform. Dyn.* **16**, 55-74 (2009).
- [14] M. Hayashi, Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding, *Phys. Rev. A* **76**, 062301 (2007).
- [15] H. Nagaoka and M. Hayashi, An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses, *IEEE Trans. Inform. Theory* **53** (2007), pp.534-549.
- [16] C.W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [17] F. Hiai, M. Ohya and M. Tsukada, Sufficiency and relative entropy in $*$ -algebras with applications in quantum systems, *Pacific J. Math.* **107**, 117-140 (1983).

- [18] F. Hiai and D. Petz, The proper formula for relative entropy and its asymptotics in quantum probability, *Commun. Math. Phys.* **143**, 99-114 (1991).
- [19] A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
- [20] E.L. Lehmann and J.P. Romano, *Testing Statistical Hypotheses*, 3rd ed., Springer, New York, (2005).
- [21] H. Nagaoka, The converse part of the theorem for quantum Hoeffding bound, (2006), arXiv:quant-ph/0611289.
- [22] M. Nielsen, and I. Chuang, *Quantum Computation and Quantum Information*, (Cambridge Univ. Press, 2000).
- [23] M. Nussbaum and A. Szkola, The Chernoff lower bound for symmetric quantum hypothesis testing, *Ann. Statist.* **37** (2009), 1040-1057.
- [24] T. Ogawa and M. Hayashi, On error exponents in quantum hypothesis testing, *IEEE Trans. Inform. Theory* **50**, 1368-1372 (2004)
- [25] T. Ogawa and H. Nagaoka, Strong converse and Stein's lemma in quantum hypothesis testing, *IEEE Trans. Inform. Theory* **46**, 2428-2433 (2000).
- [26] I. Ojima, A unified scheme for generalized sectors based on selection criteria – Order parameters of symmetries and of thermality and physical meanings of adjunctions –, *Open Sys. Inform. Dyn.* **10**, 235-279 (2003).
- [27] I. Ojima, “Micro-Macro Duality in Quantum Physics”, pp.143-161 in *Proc. Intern. Conf. on Stochastic Analysis, Classical and Quantum* (World Scientific, 2005), arXiv:math-ph/0502038.
- [28] I. Ojima, Micro-Macro Duality and Space-Time Emergence, *Proc. Intern. Conf. “Advances in Quantum Theory”*, 197–206 (2011).
- [29] I. Ojima and K. Okamura, Large Deviation Strategy for Inverse Problem, (2011), arXiv:quant-ph/1101.3690.
- [30] I. Ojima, K. Okamura and H. Saigo, in preparation.
- [31] M. Ozawa, Quantum measuring processes of continuous observables, *J. Math. Phys.* **25**, 79-87 (1984).
- [32] M. Ozawa, Quantum perfect correlations, *Ann. Phys. (N.Y.)* **321**, 744-769 (2006).
- [33] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, (Springer, Berlin, 1993).
- [34] D. Petz, Quasi-entropies for states of a von Neumann algebra, *Publ. RIMS, Kyoto Univ.* **21** (1985), 787-800.
- [35] D. Petz, A variational expression for the relative entropy, *Commun. Math. Phys.* **114**, 345-349 (1988).
- [36] D. Petz, First steps towards a Donsker and Varadhan theory in operator algebras; in *Quantum Probability and Applications V*, *Lec. Notes in Math.* **1442** (1990), 311-319.

- [37] D. Petz, G.A. Raggio and A. Verbeure, Asymptotics of Varadhan-Type and the Gibbs Variational Principle, *Commun. Math. Phys.* **121**, 271-282 (1989).
- [38] M. Takesaki, *Theory of Operator Algebras I*, (Springer, 1979).
- [39] M. Takesaki, *Theory of Operator Algebras II*, (Springer, 2002).
- [40] F. Tanaka, Generalized Bayesian predictive density operator; In *Proceedings of 14th Quantum Information Technology Symposium(QIT14)*, Tokyo, Japan, May 29-30, 2006, 107-110.
- [41] F. Tanaka and F. Komaki, Bayesian predictive density operators for exchangeable quantum-statistical models, *Phys. Rev. A* **71**, 052323 (2005).
- [42] A. Uhlmann, Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory, *Commun. Math. Phys.* **54** (1977), 21-32.
- [43] K. Usami, Y. Nambu, Y. Tsuda, K. Matsumoto and K. Nakamura, Accuracy of quantum-state estimation utilizing Akaike's information criterion, *Phys. Rev. A* **68**, 022314 (2003).
- [44] S. Watanabe, *Algebraic geometry and statistical learning theory*, (Cambridge University Press, 2009).
- [45] S. Watanabe, Asymptotic learning curve and renormalizable condition in statistical learning theory, *J. Phys.: Conf. Ser.* **233** (2010) 012014.
- [46] S. Watanabe, Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory, *J. of Mach. Lear. Res.* **11**, 3571-3591, (2010).
- [47] J.O.S. Yin and S.J. van Enk, Information criteria for efficient quantum state estimation, *Phys. Rev. A* **83**, 062110 (2011).